DOCUMENT RESUME

ED 412 240 TM 027 484

AUTHOR Villaescusa, Tangie K.; Franklin, Jennifer; Aleamoni, Lawrence M.

TITLE Improving the Interpretation and Use of Student Ratings: A (Pilot) Training Approach.
PUB DATE 1997-03-00
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Decision Making; *Evaluation Utilization; Formative Evaluation; Higher Education; Pilot Projects; Sample Size; *Student Evaluation of Teacher Performance; *Teacher Improvement; *Teachers; Test Interpretation; *Training

ABSTRACT

Few research studies have examined the use of student ratings of instructors from the standpoint of their ability to be interpreted and their subsequent usefulness. This study examined the effects of a training session conducted by experts in the field of student ratings of instruction. The first question was to determine the knowledge, skills, and attitudes that the users of student ratings possessed, and the second was to see if training helped them use student ratings. Data was collected from 68 participants at a workshop on evaluation. Participants were faculty and administrators of institutions of higher learning. Participants completed, before and after the workshop, a revised form of the Using Student Ratings of Instruction questionnaire (J. Franklin and M. Theall, 1989). Of these participants, 59% reported never having had assistance in interpreting student ratings data and indicated interest in having such assistance. The remainder had received assistance at least once. Thirty-four percent of respondents reported never having selected, adapted, or written a student ratings form themselves, but 66% had devised such an instrument at some point. Fifty percent of respondents reported never using student ratings for personnel decision-making purposes. The small sample size discouraged researchers from examining questions of predicting use from knowledge and attitudes about student ratings or from examining constructs inherent in the questionnaire. Additional studies are planned. (Contains six references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 412 240

Improving the Interpretation and Use of Student Ratings: A (Pilot) Training Approach

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Tangie Villaescusa
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Tangie K. Villaescusa
University of Arizona

Jennifer Franklin
University of Arizona

Lawrence M. Aleamoni
University of Arizona

TM027484

Paper presented at the American Educational Research Association's Annual Meeting in Chicago, Illinois, March 24-28, 1997. The authors would like to thank the Center for Educational Development and Assessment (CEDA) for their cooperation and support in the data collection process.

BEST COPY AVAILABLE

IMPROVING THE USE AND INTERPRETATION OF STUDENT RATINGS: A (PILOT) TRAINING APPROACH

Background

Through the last seventy years of research, the mystery of *why* we use student ratings of instruction has been solved: Their intended use in the 1940s, as it is now, was for students to be afforded the opportunity to rate the effectiveness of the instruction and instructional delivery offered them. When provided the results of student's ratings data, professors are expected to use the ratings not only a measure of their teaching effectiveness, but more importantly as a tool for the improvement of their teaching practices. Administrators receiving these data, are expected to use them as a part of the institution's decision making process for promotion and tenure purposes. Students previewing these results, are expected to make curriculum decisions regarding their courses of study.

However, given the plethora of research on the topic of student ratings, the question that remains virtually unanswered is that in order to be a savvy consumer of these rating's results, *what* must one know? What do any of these users of student ratings really know about interpreting the results that they receive in order to make them useful? What must one know in order to effectively use student ratings data? More importantly, how does one go about obtaining the information necessary to accurately interpret student rating's data?

Surprisingly few studies have examined the use of student ratings from the standpoint of their ability to be interpreted and hence their usefulness. Franklin and Theall (1989) acknowledged the literature suggesting the way users of student ratings interpret and/or use the data that they are presented is almost non-existent. Results from the 1989 Franklin and Theall study suggest that although one-third of their respondents indicated that they had used ratings results as part of promotion and tenure decisions, 50% of that subgroup were not able to correctly answer the most important knowledge questions about the ratings presented to them. However, those reporting the receipt of assistance in interpreting their ratings, scored significantly higher on the overall number of correct knowledge items than those who had not received any assistance. Not surprisingly, in most cases, those with the best attitude towards ratings were also likely to be the most knowledgeable about their rating's results.

Most studies of the usefulness of student ratings have focused on the combining of rating's results with expert consultation (Aleamoni, 1978; McKeachie et al., 1980; Cohen, 1980). Although findings have clearly suggested that individual faculty can improve their instructional skills, we are somewhat skeptical of the feasibility of such suggestions, based upon the limited number of available consultants and the scarce resources available today in higher education. Moreover, there appears to be no data suggesting that administrators or working committees (e.g. merit review, promotion and tenure) routinely consult student rating's experts when interpreting results for use in the promotion and tenure decision making process.

Administrator's involved in these decision making processes, appear to have at least three alternatives available to them in order to become more knowledgeable about the interpretation and utilization of student rating's results, where promotion and tenure decision

making are involved. The first alternative is the consultation of the pertinent literature. However, according to the findings of Franklin and Theall (1989) this literature consultation approach did not appear to be a widely utilized option. Working with consultants, or experts in the field is a possible second alternative. The literature (McKeachie, 1987; Wilson, 1987) makes mention of this alternative and in fact even supports its use (at least on experimental basis), but the actual practice may not occur with great frequency. And finally, these administrators and decision makers can attend and participate in interpretive workshops held by expert consultants that are geared toward the dissemination of important content knowledge of student rating's data. No evidence in the literature suggests that this strategy has been systematically examined.

The importance of this study to the field

This study provides an examination of the effects of a training session conducted by experts in the field of student ratings of instruction. Training, whereby decision makers would be afforded the opportunity to learn basic assumptions of student ratings interpretation, including the introduction of student rating's myths. Such training, if found to make a difference, could prove a cost effective approach to an obviously existing problem: users of ratings data being unaware of prudent practices of interpretation and use. The effects of such indiscriminate practices must be far reaching, but unfortunately data is not available suggesting exactly how devastating such misuse can be.

This study adds to the existing body of literature to the extent that no endeavor such as a training session has been examined systematically. Moreover, the findings might suggest that the inclusion of such a training program could provide stakeholders a viable alternative to the current strategies for using and interpreting student ratings data; currently considered to be faulty at best (Franklin & Theall, 1989).

Goals of the study

The purposes of this study are two-fold:
First, to assess the knowledge, skills and attitudes of the users of student ratings of instruction (via a questionnaire designed to elicit the necessary data), and second, to examine how these attributes are affected by expert training. Specifically, the research questions are:

1 What knowledge, skills and attitudes about student ratings results do the users of these data possess?

2. Does training help aid users of student ratings in their interpretation and perceived use of the rating's data?

Participants

Data for this study was collected from 68 participants who had attended a workshop entitled "Developing a Comprehensive Evaluation System". The four-day workshop was held in St. Louis, Missouri, in October 1996 . The participants were faculty and administrators of institutes of higher learning from throughout the continental United States and Puerto Rico.

The Workshop: An Overview

The four day workshop is facilitated by two experts in the fields of comprehensive faculty evaluation systems and student rating's of instruction. The primary goal of the workshop was to provide its participants with step by step procedures to use in the development of a comprehensive faculty evaluation system that specifically delineates the role Peers, Students, Administrators and others play in the evaluative process. Additionally, the facilitators addressed the issues of evaluating teaching, research and service; the use of peer and student ratings; techniques for combining data from multiple sources; and the development of questionnaires, forms, and operational policies. Moreover, methods for using evaluation information as feedback for instructional improvement, as well as for promotion, tenure and merit pay decisions were also addressed.

The Workshop: An Agenda

The first two days of training included topics such as: "The Portfolio System for Gathering and Maintaining Evaluative Information", "Determining the Faculty Role Model", "Determining Component Weights",and "Application to Personnel Decisions". The following two days focused on topics such as "Student Ratings: Myths Versus Research Facts", "Peer Evaluation", "Differential Relationships of Student, Instructor, and Course Characteristics to General and Specific Items on a Course Evaluation Questionnaire", "Techniques for Designing a Course Instructor Rating Form" and "Evaluating Academic Administrators".

The Instrument

The participants were asked to anonymously complete, pre and post, a revised form of the "Using Student Rating's of Instruction" questionnaire. This instrument was originally designed by Franklin and Theall (1989) to measure the-knowledge and attitudes of users of student ratings data. Additionally, the researchers sought to explore the perceptions of these users in terms of their skills, and hence, application of such skills. The questionnaire, in its current form, is comprised of three components, the first containing 54 statements; 51 of which were devised to measure basic, global knowledge of and attitude's toward student ratings. The questionnaire provides a 5-option Likert response scale that includes Strongly Agree, Agree, Uncertain, Disagree and Strongly Disagree. The remainder of the statements in the first section were incorporated to summon demographic data. In the second section, the respondents were supplied with a set of "dummy data" and given the opportunity to apply their student rating's

skills to 12 situational statements. In the third and final section, the participants are asked three additional questions: the role that they play in selecting student rating's forms at their institutions, how often they use rating's results in making personnel decisions, and finally, if they have ever received assistance from an expert in such endeavors.

The Procedure

On day three of the workshop, prior to the beginning of instruction on student ratings, the participants were given the revised form of the "Using Student Ratings of Instruction" questionnaire by one of the workshop's facilitators. On day four, following the student rating's instruction and prior to the respondent's departure, they were once again asked to complete the "Using Student Ratings of Instruction" questionnaire. For those unable to complete the questionnaire prior to their departure, they were given a self-addressed, stamped envelope along with the post questionnaire and were asked to complete and mail the questionnaire within two days.

The Results

Sample Demographics

1. Fifty-eight percent of the respondents were female, forty-two percent of the respondents were male.

2. While 3% reported never having had assistance in interpreting student ratings data and were *not* interested in having assistance, 59% reported never having had assistance in interpreting student ratings data and were interested in having assistance. The remainder had received assistance in interpreting student ratings data at least once.

3. Thirty-four percent of the respondents reported never having selected, adapted or written a student ratings form for either their own use in improving instruction, their own use in personnel decision making, or for the use of others in either improving instruction or for use in personnel decision making.. Sixty-six percent reported having selected, adapted or written a student ratings form for either their own use in improving instruction, their own use in personnel decision making, or for the use of others in either improving instruction or for use in personnel decision making.

4. Fifty percent of the respondents reported never having used student ratings for personnel decision making purposes. The other fifty percent reported having either used student ratings in the past or were currently using student ratings for personnel decision making purposes

The Analysis

A reliability analysis of the questionnaire focused on its subscales. The obtained alphas were as following:

The **Knowledge subscale** yielded an alpha of **.81**.
The **Attitude/Opinion subscale** yielded an alpha of **.64**.
The **Use subscale** yielded an alpha of **.60**.

Group Differences

The **Knowledge** subscale yielded an obtained t score of *4.58, n=27, p=.000.
The **Attitude/Opinion** subscale yielded an obtained t score of *2.11, n=27, p=.044.
The **Use** subscale was only analyzed using frequency scores because of a lack of responses; presumably due to incorrect information being provided the participants in the vignette.

***Using a correlated group's t test, significance was found in both the knowledge and attitudinal subscales.**

Additional Analyses

The researchers had planned to apply factor analytic techniques to the data, in an attempt to explain the variance and identify hypothetical constructs inherent in the questionnaire. Additionally, predicting **use** from **knowledge** and **attitudes** was also planned via a series of simultaneous regression equations. However, with the small sample size they have decided not to pursuit these endeavors until more field testing of the instrument can be done and a greater sample size can be obtained.

Discussion/Conclusions

Although significance was found in this pilot study, such critical mysteries regarding knowledge of users within the field of student ratings, simply cannot remain unsolved. If student ratings are to be used effectively, then we as researchers must continue in our pursuit for knowledge on the use of their results. In order for us to promulgate the utilization of student ratings, we must not only discover what the users know, but more importantly, how they apply their knowledge of ratings. The bottom line is (as Franklin and Theall eluded to in 1989) a valid and reliable rating's instrument can only be as valid and reliable as the users who use them. When you stop to think about it, isn't the **use** of rating's data the most fundamental aspect of their existence?

Moreover, continued studies of a training approach may be as important and necessary an endeavor in the improvement of knowledge and attitudes among faculty members and administrators, as has been just about anything else in recent years. This continued research just may shed light on some very integral questions that still remain unanswered by the community of scholars that study student rating's research.

Limitations of the Study

1. No random assignment was used, which could have aided in the attempt to control for alternative explanations.
2. No Control Group was used, which could have provided a stronger baseline for evaluating the effects of the workshop.
3. The time elapse for some participants in returning post questionnaire was longer than others. This might suggest that something other than the treatment was responsible for the posttest data.
4. Sampling error that could have occurred as a result of a small sample size (and perhaps lack of variability).

References

Aleamoni, L. M. (1978). The usefulness of student evaluations in improving college teaching. Instructional Science, 7, 95-105.

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. Research in Higher Education, 13, 321-341.

Franklin, J. & Theall, M. (1989). Who read ratings: Knowledge, attitude and practice of users of student ratings of instruction. Paper presented at the annual meeting of the American Education Research Association, San Francisco, California.

McKeachie, W., Lin Y., Daugherty, M., Moffett, M., Neigler, C., Nork, J., Walz M. & Baldwin, R. (1980). Using student ratings and consultation to improve instruction. British Journal of Educational Psychology, 50, 168-174.

McKeachie, W.J. (1987). Can evaluating instruction improve teaching? New Directions for Teaching and Learning, 31, 3-8.

Wilson, R. C. (1987). Toward excellence in teaching. New Directions for Teaching and Learning, 31, 9-24.